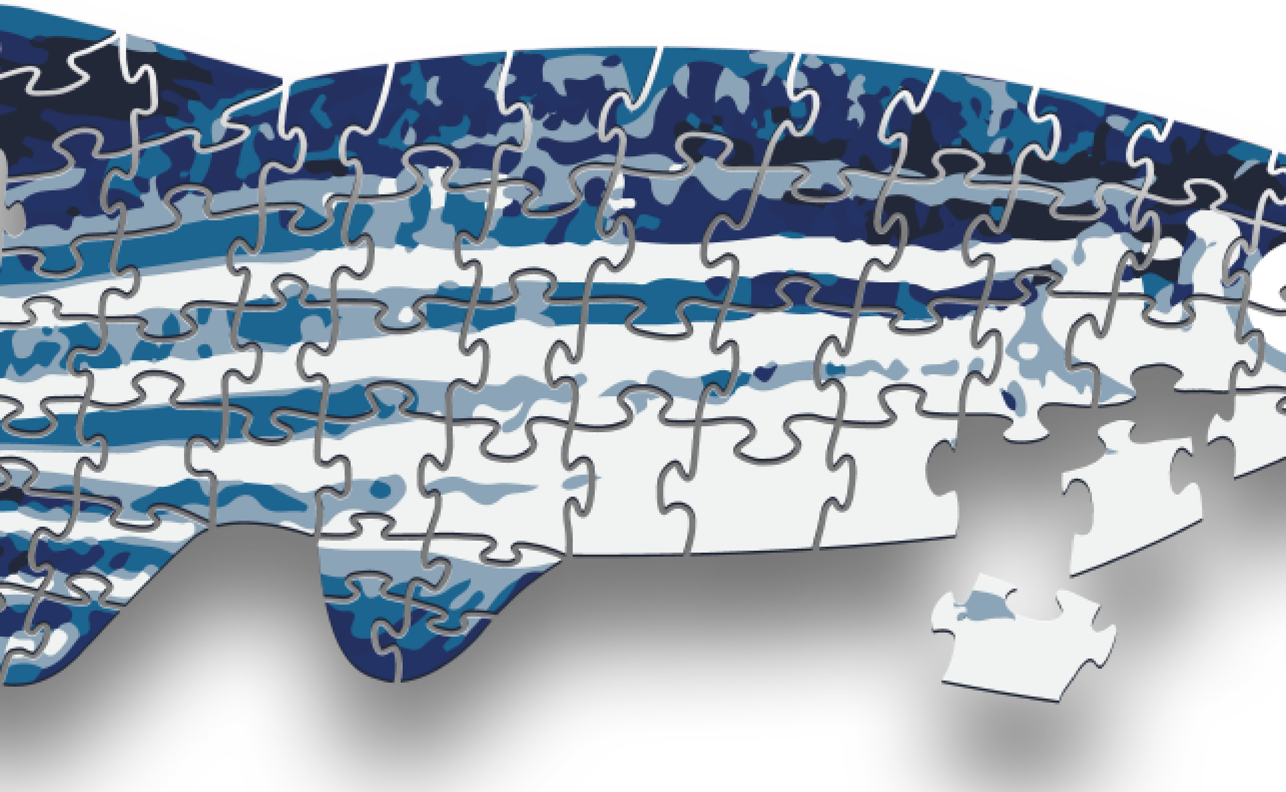
# Advancing zebrafish models in proteomics

Duivesteijn, S.J.

## Citation

Duivesteijn, S. J. (2018, October 30). *Advancing zebrafish models in proteomics*. Retrieved from https://hdl.handle.net/1887/66790

**Author**: Duivesteijn, S.J.
**Title**: Advancing zebrafish models in proteomics
**Issue Date**: 2018-10-30



# Chapter 5

# Zebrafish Models in Proteomics

**S.J. van der Plas-Duivesteijn, E.L. Benard, I.C.R.M. Kolder, K.A. Stam, A.H. Meijer, M. Palmblad**

Ready for submission

## ABSTRACT

Zebrafish models have been shown to be powerful tools in genomics studies for drug discovery and drug development and their application in biomedical research is increasing. The use of this animal model in the field of proteomics is also accumulating, although it is still less frequently used within this omics field than in the field of genomics. In this review we provide an overview of the possibilities, limitations and practical considerations particular to this animal model in proteomics, and evaluate available tools and resources suitable for zebrafish proteomics.

## INTROCUCTION

The zebrafish (*Danio rerio*) is a tropical fish originating from Southeast Asia, where it lives in brackish waters, and is a member of the Cyprinidae family. It was introduced as an animal model in the 1970s (1). The use of the zebrafish as an animal model in biomedical research has emerged over time (2-6). Currently, mutant strains and zebrafish cell lines are available (7). Clearly there are major differences between the anatomy of human and zebrafish, reflecting their evolutionary distance. Nevertheless, despite their taxonomic distance, approximately 70% of their protein-coding genes shows strong similarities (8). In addition, amino acid sequences of proteins are evolutionary conserved between mammals and zebrafish (9-11). Also, several components of the main immune signalling pathways (e.g. Toll like receptors, and Myeloid differentiation primary response gene 88) are strongly conserved in both fish and mammals (5, 12-15). Summarized, these studies indicate the homology of the zebrafish animal model to mammals on anatomical, physiological and molecular levels, and highlight the applicability of the zebrafish animal model in biomedical research.

Many studies that make use of the zebrafish animal model are focussed on genomics and transcriptomics. Gaining a more complete insight on e.g. immune responses to infection and drug treatment, requires a multiple omics approach (including genomics, transcriptomics, proteomics, and metabolomics) (16-18). Proteomics, the study of all proteins of an organism or part of an organism, is a growing field of research. Proteomics is important because it describes direct bio-functional signalling, reflecting the operating machinery for nearly all biological functions. Mapping proteins and protein interaction, will therefore contribute to the clarification of biological pathways followed by the understanding of diseases and development of possible treatments.

A quick bibliometric search of zebrafish proteomics clearly shows that the model system is emerging in this field, with the number of publications rising from 9 in 2001 to 245 in 2017 (figure 1). Launching the first zebrafish spectral library (2014), was an important addition toward the maturation of this model species within proteomics (19). The experimental knowledge on the application of the zebrafish animal model in proteomics, is summarized here by reviewing the possibilities, limitations and practical considerations of the model system. In addition, we evaluate available tools and resources suitable for zebrafish proteomics.

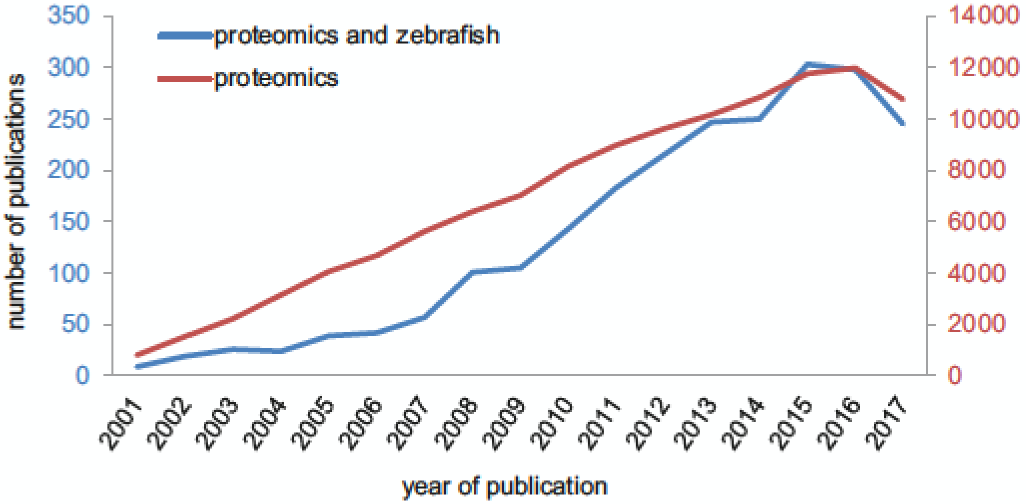


**Figure 1: Plot indicating the increase in the number of publications** of proteomics studies applying the zebrafish animal model (left axis, blue), compared to the increase of proteomics studies in general (right axis, red) for terms "proteomics AND zebrafish" and "proteomics", respectively. The dip in 2017 is caused by a delay between publication and full indexing and availability in PubMed. Literature search using: https://europepmc.org , on April 9, 2018, using the search between publication and full indexing and availability in PubMed.

## THE ZEBRAFISH AS AN ANIMAL MODEL

Multiple animal models are available, each with their own advantages and disadvantages (2). The zebrafish animal model is widely used and accepted within human disease studies, including drug discovery and development. Even though anatomically different to human, the zebrafish animal model shows a close genetic relation to human (often represented in phylogenetics) (8) with similarities to several components of immune signalling pathways (5, 12-15), demonstrating its applicability to such studies. As in mouse studies, many strains have been developed for the zebrafish animal model (7, 20). Like the human population, wild type zebrafish strains contain high numbers of polymorphisms, which make the zebrafish animal model extremely relevant to apply in human health studies (21, 22).

An experimental design to study disease mostly includes the early life stages (embryos and larvae) of the zebrafish, which enable the application of high throughput screening (HTS). The natural transparency of the embryos allows the use of fluorescence techniques to follow-up on infection *in vivo* (23-30). In addition, innate immunity (functioning from 24 hours post fertilization) and acquired immunity (not active during the first weeks, and matures at 4-6 weeks) can be studied separately (31), which is a powerful fool tool to study immune reactions in drug discovery and development. Due to the advantages of transparency, zebrafish strains that remain transparent as adults are now also available (32). In order to meet biomedical study conditions, various mutant strains and zebrafish cell lines have been created (7, 33).

The use of animal models in a laboratory setting is ethically restricted to certain rules. The reflection on reduction (minimise animal use), replacement (methods, strategies and approaches not making use of animal models), and refinement (improvement of procedures and welfare to minimise pain, suffering, distress or lasting harm) should be applied to each study using animal models (34). This widely accepted reflection is meant to protect animal health and welfare, and should exactly describe sedation of and humane endpoints for the animals. Unfortunately directives and policy on the protection of animals used for scientific purposes show large variations between national and regional authorities with respect to application and endpoint (35). According to the EU Directive 2010/63/EU, early life-stages of zebrafish are not protected as animals until the stage where it is capable to feed independently (for zebrafish at 5 days/120 hours post fertilization) (36). This means ethical approval is not required for working with zebrafish embryos and larvae until 5 dpf in scientific studies (another practical advantage when applying embryos/larvae compared to the use of adult stages within a study). The defined stage of independent feeding is influenced by temperature, local conditions, and genotype, causing variation in the defined stage from 5-6 dpf. These factors should be considered to meet the ethical rules that protect this animal and proper staging of larvae 5 dpf should follow the morphological criteria by Kimmel *et al.* (37). Directives and policy are constantly moving, and should therefore be taken into account as well when applying animal models in scientific studies.

## PRACTICAL CONSIDERATIONS IN A PROTEOMICS EXPERIMENTAL DESIGN

As mentioned before the zebrafish can be divided in two main categories as an animal model: adult and developmental stages (embryo/early larvae). Adults are preferably kept in groups, where embryos can be separated, e.g. in 96 well plates. In addition, zebrafish adults produce approximately 200 transparent eggs per pair per week, and development progresses fast, while breeding fish until adult stage takes months. This indicates that embryos are more suitable for high throughput processing and screening than adults. High-throughput screening (HTS) is an important method used in human drug discovery and development. The increased use of zebrafish as an animal model and its applicability for HTS has led to the development of several laboratory techniques. For example, a spawning system, named Ispawn, was developed to scale up zebrafish breeding and easily collect eggs (38). A robot injector automated infections of zebrafish embryos, and the additional developed COPAS XL (Union Biometrica) automatically sorts diseased embryos/larvae based on the illuminance of fluorescent signals that can be labelled to the source of infection (e.g. fluorescent bacteria or cancer cells) (23, 28). In addition, several routes of injections are well established for zebrafish (39). Moreover, techniques using fluorescently labelled cells (e.g. macrophages or neutrophils) assist in visualizing the effect on a cellular level *in vivo* without sacrificing the organism. (30, 40). Furthermore, genomic and transcriptomic screens can be applied to the model, providing an overview of most genes active or inactive at that moment, often visualized in heat maps. In combination these techniques make the zebrafish animal model suitable for HTS (41). Within proteomics, a total protein map, visualized in something like a proteomics heat map, would be desirable to level with genomic and transcriptomic studies. The above mentioned techniques highlight the possibility to apply the zebrafish animal model in HTS, but what about its application to proteomics at present? We therefore review the practical considerations of the zebrafish animal model and its use in proteomics step by step.

### Sample preparation

Sample preparation is a first step in proteomics. Where adult zebrafish can be cleared from water, larvae and embryos cannot be processed without a little excess fluid which can interfere in the proteomics extraction process. In addition, zebrafish embryos require additional dechorionation and deyolking steps prior to protein extraction.

The embryonic chorion can easily be digested using pronase. A solution of 2 mg/ml pronase (Roche Applied Science) in egg water (42, 43) can be applied to large badges of embryos at once in a petri dish. Embryos younger than the tailbud stage require coated dishes, since they stick to the plastic of the petri dish. A disadvantage of pronase is its protease activity, which is not desired when applying proteomics, and therefore requires several washing steps to remove (43). Even though robotic chorion removal is available (44), a technique that makes use of 96 well plates, used to expose and remove fluids in HTS techniques like the fish and chips microfluidic perfusion platform (45, 46), would be more desirable for HTS.

For nutrition, zebrafish embryos and early larvae rely on a yolk sac throughout development until exogenous feeding begins (5-6 days post fertilization, depending on water temperature, local conditions, and genotyping). The high abundance of yolk proteins (mainly consisting of vitellogenins) seems a major obstacle when applying proteomics, due to its interference with any proteomic application (43, 47). To enable the detection of less abundant proteins, it was suggested to apply deyolking, a procedure to manually separate and remove the yolk from the body, prior to protein extraction (43, 48). A later study on the impact of deyolking, described that deyolking depleted molecular pathways associated with calcium signaling and immune response signaling (49). It was therefore advised to use both yolk-intact and deyolked embryos to maximize protein coverage for shotgun proteomics studies (49). Furthermore, deyolking becomes more difficult throughout development, since the surrounding tissues are becoming more rigid (S.J. van der Plas - Duivesteijn, personal communication). In addition, deyolking does not result in complete yolk removal, leaving the remaining abundant yolk proteins to interfere the detection of less abundant proteins. According to our own experience, we mark deyolking as a time consuming procedure which has a low accuracy. With the development of multiple omics and drug screens in mind, which require HTS, we would advise to invest on developing automated and faster deyolking methods. For example developing antibody coated LC-columns that extract vitellogenins proteins in an automated way e.g. like IgY RuBisCO coated columns to remove Ribulose-1,5-bisphosphate carboxylase/oxygenase, the most abundant protein in plants (50).

When applying zebrafish embryos into an experimental design, their natural development should be considered carefully. Within 5 days a fertilized zebrafish embryo develops from a fertilized egg into a free swimming larva (37). During this process, developmental variability can occur (51), which can be caused by several parameters (52). When separating proteins from different embryonic/larval stages using SDS-PAGE gel, the magnitude of changes between the proteomes becomes clear (figure 2), indicating the importance of developmental synchronization, which was also addressed earlier (4). Even though the use of embryos from the same lay of a single pair of adults will be synchronous enough for most studies, with fast developing inventions at present, precise staging of embryos might become a robotic procedure (using automated image recognition) in the near future to serve advanced HTS.

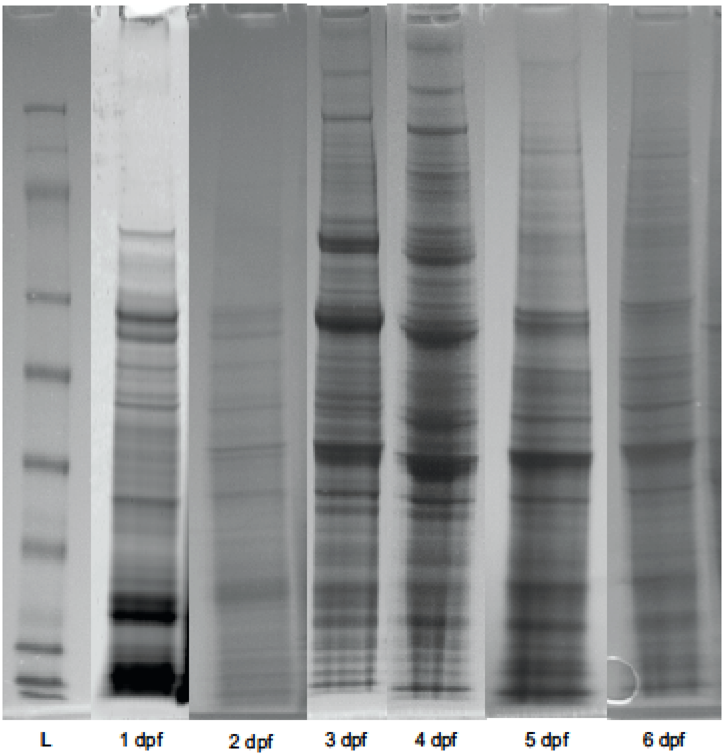


**Figure 2: Zebrafish embryo/larva proteome at different developmental stages (1-6 dpf)** visualized using SDS-PAGE gel. Figure produced from SDS-PAGE separation used to build the zebrafish spectral library (19).

### Protein extraction

A second step in proteomics is protein extraction. Even though essential, a (commercial available) protein extraction kit, validated for the zebrafish animal model, is still lacking. Currently, protein extraction of zebrafish is based on methods described in literature, which have not been validated in a comparative experiment. We here review the different methods found in literature. For homogenization, an essential step in protein extraction, adult zebrafish can be grinded to powder by using liquid nitrogen. Methods applied to homogenize embryos are sonication (48), freeze thaw cycles (49) or making use of the bullet blender (19). Contrarily it was mentioned that early zebrafish embryos will dissociate rapidly in protein extraction buffer (43).

One advantage of using the bullet blender to homogenize is that it does not require prior dechorionation and it can be applied to adult tissues (more rigid than embryonic tissue) as well (19, 53). In addition, the bullet blender is able to process 24 eppendorf tubes in one automated run, and is a closed system, preventing human keratin contamination. Unfortunately the bullet blender cannot process entire zebrafish adults. Contrary, the use of liquid nitrogen and micropestles increase the risk of keratin contamination to the samples and increases the risk of cold burn to the scientist, and moreover, is a time consuming and manual procedure (which often result in lower accuracy compared to automated procedures).

Different protein extraction buffers used to extract proteins from homogenized samples are described in literature. The use of different protein extraction buffers and methods can lead to a variety in study results. Where one method shows abundancy in cell membrane proteins, the other will show a higher abundance in nuclear proteins (48). Therefore sequential fractionation is advised (54). Nevertheless this method does not emphasize HTS.

Another consideration is the addition of protease inhibitors to the protein extraction buffer (55). Even though not explicitly described in literature, adding protease inhibitors to the protein extraction buffer has a positive effect when applied to study zebrafish proteomics (S.J. van der Plas – Duivesteijn, personal communication, (19, 43)).

When applying a shotgun bottom-up approach within proteomics, another challenge is the concentration of proteins that can be extracted from a sample. The experimental design for the creation of a zebrafish spectral library to cover deeper protein coverage, described the use of more than one organ to reach sufficient protein concentration levels (19), while for other experimental designs a single embryo is sufficient (48). In addition, the amount of embryos required for cell dissociation, to study protein effects *in vivo* on cellular level, limits its use to be applied to shotgun bottom-up approach studies (S.J. van der Plas – Duivesteijn, personal communication).
In summary, a validated commercially available protein extraction kit for the zebrafish animal model will hopefully soon be available, to facilitate the use of this animal model in proteomics screening.

Protein extraction is followed by liquid chromatography (LC) separation of the peptide mixture, or SDS fractionation and thereafter LC separation. The last laboratory based step is the detection of the peptides by mass spectrometry. Methods applied to protein mixtures and mass spectrometer settings are well described in literature and do not require specific alterations for its use on zebrafish material, and therefore is not further described in this review.

### Data analysis

Finally in these practical considerations, we review proteomics data analysis. Many different software tools are available to match tandem mass spectra to peptides, identifying the peptide and infer the protein (56).

For a standard bottom-up shotgun proteomics experiment, peptide identification can be achieved by at least three different methods: *de novo* sequencing, searching against sequence databases, and matching against real spectra in spectral libraries. *De novo* sequencing does not require any a priori sequence information from the species, but is the least sensitive method of these three. Database searches correlates acquired $MS^2$ spectra with *in silico* generated spectra from a protein sequence database, and works best for completely sequenced organisms, although false-discovery rates can be estimated for partially sequenced ones as well (57). Peptide identification using spectral libraries are based on the correlation between query spectra and previously acquired spectra. This is the most sensitive method, but requires both the sequence database and a diverse spectral library. The search space when using spectral libraries is typically smaller than in database searches, and faster when searching large datasets, especially when allowing for variable post-translational modifications (58, 59). In 2014, we released a spectral library for zebrafish (19). As is the case with the spectral libraries available for other model species, the library can be extended by inclusion of data from additional proteomics experiments, or rebuilt using data from more recent instruments. To be able to detect novel proteins, peptides or modifications, a combination of database and library search methods is advised (19). This can be practically accomplished in several freely available tools, such as the Trans-Proteomic Pipeline (60), using Comet for the sequence database search, SpectraST for the spectral library search, PeptideProphet to validate the peptide-spectrum matches, and iProphet to integrate the database and library search results.

After peptide identifications are aggregated and proteins inferred and quantified, zebrafish proteomics data can be further analysed and presented much like quantitative proteomics data from any other organism, including gene ontology and pathway enrichment analysis without necessitating conversion to orthologs from a different species. For example, there are currently (April 2018) 1,560 *Danio rerio* pathways in Reactome, which is comparable with mouse (1,603 pathways). The Comparative GO web tool was recently updated (61) and extended for zebrafish omics data. Though many of these tools were developed for genomics data, they work equally well for proteomics data, as proteomics data is typically gene-centric. Proteins are identified by association with a particular gene and each protein identification (using the UniProt FASTA file or our spectral library) also contains the gene name. This means the quantitative proteomics data can be directly plugged into these and many other tools for GO and pathway enrichment analysis and visualization.

## PROTEOME PHYLOGENETICS SUGGESTS APPLICATION OF COMPLEMENTARY MODELS

Shared ancestry and similarities between humans and animal models are crucial in hypothesis determination, and therefore important to apply when designing experiments for human disease and drug development. The most currently used methods to study molecular evolution/phylogenetics, make use of DNA sequences (62). However, before high-throughput DNA sequencing methods were developed, patterns of proteins or enzymatically cleaved peptides (which were separated by chromatographic or electrophoretic methods) were used (63, 64). Recent proteomic studies that applied phylogenetics, showed that, by using stable and abundant proteins, it is possible to analyse much older specimens than with DNA-based methods (65-69). Additionally, proteomic phylogeny demonstrated new discoveries in evolutionary associations (70). The comparison between DNA and protein is not trivial, since one gene, via mRNA processing and post-translational modification (PTM) of the protein sequence, can give rise to more than one *proteoform* (and thus finally to more than one biological function). This in turn may affect the distance as measured by shared peptide or protein features, independent of the full DNA sequence for the same protein. A recent study explored phylogenetic relationships of primates, using a purely proteomics approach, with an algorithm that directly compares tryptic peptide tandem mass spectra between species (71). This method, based on a common bottom-up or "shotgun" proteomics method, may also be applicable in the design of proteomics experiments, and can be informative in choosing an animal model system.

In proteomics studying the frequency of shared tryptic peptides between organisms is of great importance, as these can cause interferences in some experiments while being beneficial in other studies. For example, shared peptides synthetized as stable isotope labelled standards can possibly be used for both human and animal studies, which will reduce the need to synthesize and validate such peptides for each (model) species separately. Furthermore it may be interesting to know whether spectra can be used across (model) species when identifying peptides using spectral libraries and no library for the species of interest is available.

In 2014, the first zebrafish spectral library was published (19), based on a novel approach of using dissected zebrafish organs to produce a wider proteome coverage and better spectral quality. Unfortunately, due to the small size of the zebrafish, not all organs could be easily dissected, for example it is difficult and extremely time consuming to dissect the different parts of the heart. In addition, embryos are even smaller, making it challenging to accomplish deep proteome analysis in single embryos, especially if one wants to perform protein fractionation or enrich PTMs. Moreover, human and mouse libraries contain spectra from specific cell lines as well. Even though zebrafish cell lines are available (33, 72), they were not added to the library at present. A recent study showed the applicability of highly conserved peptides, bridging the gap of the zebrafish models limitations due to its small size, by using a larger but closely related carp (*Cyprinus carpio*) (73). In addition, where zebrafish produce hundreds of eggs per female, carp lay several hundreds of thousands of eggs per female. Especially in high throughput screens, the amount of eggs produced can serve as a limitation, therefore, another advantage of using carp as a complementary model to zebrafish, is the larger production of carp eggs. This study, in which the limitations of one model are addressed by the other, stresses the importance of applying "proteome phylogenetics", especially since it was shown that applying highly conserved peptides between model species can be reapplied in related animal models (73).

## RESOURCES

Many online recourses are available assisting the application of the zebrafish animal model for research purposes (74). For the field of proteomics no resource is available focussed on the zebrafish animal model alone. Instead, proteomics resources are arranged/centralized based on this omics field explicitly. The largest information resource for proteomics is Uniprot (75). This online resource provides proteomes for species that have their complete genome sequenced, containing a reference proteome for zebrafish. Uniprot provides high-quality data of protein sequences accompanied with functional information. PeptideAtlas (76), a public accessible compendium of peptides that are identified in large tandem mass spectrometry proteomic experiments, released a first zebrafish build on April 25, 2014 (77). Unfortunately, this resource is built on data from a pending publication and unpublished results. For spectral library searching, the National Institute of Standards and Technology (NIST) develops peptide mass spectral libraries with the purpose to provide peptide reference data for disease-related biomarker discovery (78). On September 1, 2015, NIST released the first zebrafish spectral library (19, 78). Our recently published omics data visualization tool, COMICS, can generate anatomical representations of gene and protein expression, including for zebrafish (figure 3), stimulating a multiple omics approach (79). Conglomerate resources for zebrafish only are also available, but are focussed on laboratory methods, strains, cell lines, genomics, and transcriptomics (table 1) (74). Via these well-known resources by researchers applying the zebrafish animal model, notifications to protein repositories can be mentioned. On April 20, 2015, the zebrafish spectral library was mentioned which redirected to the NIST webpage via an active link. With above described proteomics resources for the zebrafish animal model, and to avoid an ocean of online webpages developed for the zebrafish animal model, we see possibilities for implementation of proteomics data within existing genomic recourses.

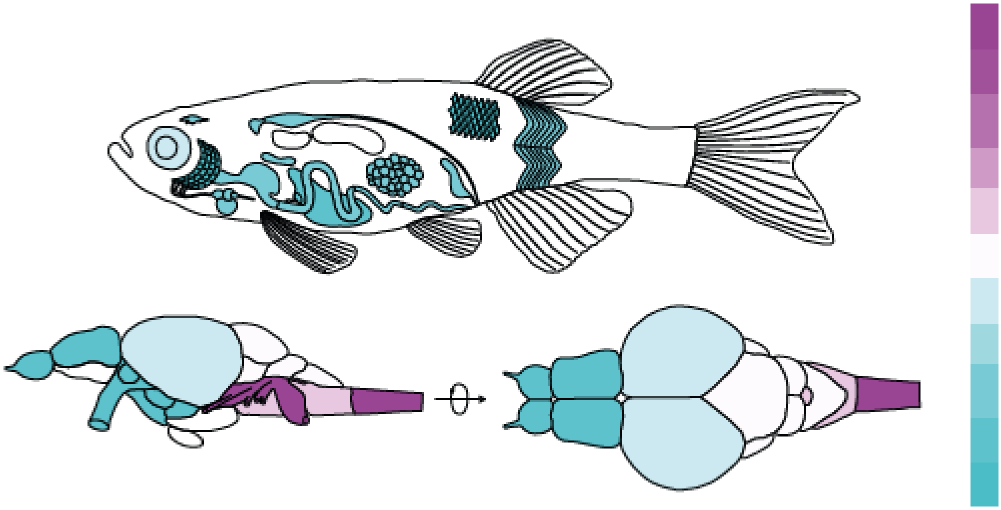


**Figure 3: Log10 ratio of recorded observations of *ca2* (cyan) and *mbpa* (magenta) gene expression** in ZFIN visualized by COMICS (79). This particular illustration primarily summarizes what information is available in the ZFIN gene expression database, but also says something about the spatial dependence of expression of these two genes. Any anatomically annotated omics data can be visualized by this software tool. The use of the ZFA anatomical ontology is critical to map the observations from multiple experiments to the same layer or level of granularity in the anatomogram.

**Table 1: Zebrafish animal model resources**

| Name of resource | Aim of the resource | Webpage of the resource |
|---|---|---|
| ZFIN | zebrafish model organism database | https://zfin.org |
| zebrafish international resource centre (ZIRC) | repository for zebrafish strains, material and research information | http://zebrafish.org |
| Zebrafish book (45) | a practical handbook for zebrafish protocols | https://zfin.org/zf_info/zfbook/zfbk.html<br>and<br>https://wiki.zfin.org/display/prot/ZFIN+Protocol+Wiki |
| The Journal of visualized experiments (JOVE) | creating and publishing videos of scientific experiments allowing scientists, educators and students to see the intricate details of cutting-edge experiments rather than read them | https://www.jove.com |

Zebrafish-specific and general resources for this animal model that are widely used among researchers in the field. ZFIN and ZIRC are sister organizations. JOVE is a repository of tutorial and practical video instructions that include videos on mass spectrometry, proteomics, data analysis as well as actual handling of the zebrafish.

## EXAMPLE: ZEBRAFISH MODEL TO STUDY TUBERCULOSIS INFECTION

The zebrafish is a well described model to study infectious diseases, and often applied to study tuberculosis (TB) infection (2, 5, 80-82). To study mycobacterial infection in the zebrafish model, the bacterium *M. marinum* (a natural host of zebrafish) is used (81-84). Genetically, *M. marinum* is closely related to *M. tuberculosis* (85, 86), and both have the the ability to replicate inside macrophages and induce the formation of granulomas, the hallmark pathological structures of TB (84, 87). The zebrafish mycobacterial infection model, using *M. marinum*, mimics many aspects of human tuberculosis infection (82, 84), and the use of this model has led to new insights into mycobacterial pathogenesis relevant to the human TB disease (88-90). With only 16 publications in 2016 (PubMed search November 2017), studies on the zebrafish TB infection model using *M. marinum* infections barely touched the field of proteomics. Nevertheless, proteomics seems a powerful tool to answer important remaining immune questions in relation to TB infection (17, 91, 92).

Injecting fluorescent labelled pathogens, e.g. *M. marinum*, into the yolk and vein of zebrafish embryos is a well described technique, resulting in easy detectable bacterial growth which eventually leads to the formation of aggregates of infected immune cells, representing the early stages of the granulomas characteristic for TB (39). Unfortunately, when applied to proteomics using the shotgun bottom-up technique, this did not result in the detection of proteins or protein groups that allowed to distinguish the infected from the non-infected zebrafish embryos significantly (S.J. van der Plas - Duivesteijn, personal communication). In addition, attempting to support a multiple omics approach, adult zebrafish infected with Mma20 and E11 *M. marinum* strains from a previously well published transcriptomics infection study (93), were examined on protein expression in a shotgun bottom-up approach (S.J. van der Plas - Duivesteijn, personal communication). Only 13 proteins were found to be significantly different between the infected and non-infected adult zebrafish in all three replicas. An additional 22 proteins were found to be significantly different between non-infected and infected individuals, but this significant difference could be found in only two out of three replicas, and therefore were marked as speculative results (S.J. van der Plas - Duivesteijn, personal communication). In comparison with previous studies where many significant differences were found on the genomic and transcriptomic level, the number of significant proteins found in this proteomics study was unexpectedly low (S.J. van der Plas – Duivesteijn, personal communication). The resulting low numbers of significantly different proteins found in this adult infection study, and the lack of a possible distinction between infected and non-infected individuals in the study with embryos, could be due to the low number of replicas used to study the proteomics effect. In contrast, the same amount of replicas used in the adult infection setup did result in sufficient genomic results (93), and the infected zebrafish embryos showed severe infection after fluorescent microscopic examination (S.J. van der Plas – Duivesteijn, personal communication). In addition, the lack of a validated protein extraction kit could also have affected these results. In particular, high levels of vitellogenins were found in the embryo infection experiment (S.J. van der Plas - Duivesteijn, personal communication), which can affect protein detection of lower abundant proteins in the same detection range, emphasizing the need for a more efficient deyolking protocol. Finally, due to the time-consuming data analysis, we experienced the need for the development of software tools assisting in (more) automated screening of the data.

Several genomics studies include additional studies on cellular level of the zebrafish animal model. We therefore also applied a standard bottom-up proteomics approach on FACS sorted cells from dissociated zebrafish embryos. Dissociating immune cell populations from zebrafish using FACS has also been described in a previous study on RNA sequencing (94). Using FACS enables deeper proteome coverage and potentially simultaneous detection of zebrafish and bacterial proteins, as these can be easily distinguished, even on the tryptic peptide level. However, FACS is a slow process and requires additional expertise and equipment, and is not yet easily applicable in high-throughput workflows.

## CONCLUSION

The zebrafish has been established as animal model in genomics and transcriptomics and now also emerges in proteomic studies. Since the zebrafish is used as a powerful model within biomedical research, there is a need for the development of HTS methodology. Different zebrafish strains and cell lines have been developed to meet biomedical hypotheses. The use of zebrafish embryos show great potential when applied to HTS. Harmonization of ethical legislation between countries would benefit the use of the zebrafish animal model and assist in faster development of HTS methodology.

From a practical point of view, key technology has been developed to increase the throughput specifically in experiments using zebrafish embryo and larval models. Examples include the ISpawn system for breeding and a fully automated injection system to automate embryonic infection (23). For proteomics applications, there is a need for the development of an automated deyolking protocol, a commercially available protein extraction kit and standardized protocols for homogenization. In addition, robust and automated analysis platforms are essential when applying proteomic HTS on the zebrafish or other animal models.

Besides the development of new methodology, research communities should investigate the possibility of applying complementary species in their studies of animal models near the boundaries of their proteomics design. Rather than developing entirely new online resources, existing ones could be expanded to handle additional model systems to avoid oceans of available resources in which researchers drown when searching for information.

Altogether, lots of work needs to be done to accomplish a successful proteomics HTS using zebrafish as animal model. However, developed techniques and procedures will also be beneficiary for other studies applying proteomics. With such a successful model system as the zebrafish and proteomics as an emerging, valuable and promising field of research, we have no doubt above suggested inventions will be developed in the near future.

## ACKNOWLEDGEMENTS

The authors wish to thank the members of the zebrafish facility team of the Institute of Biology for fish caretaking, and Dmitrii Travin and Iaroslav Popov for technical assistance with the COMICS software. SJPD and MP were supported by the Netherlands Organization for Scientific Research (NWO) Vidi grant 917.11.398 (MP).

## REFERENCES

1. Streisinger, G., Walker, C., Dower, N., Knauber, D. & Singer, F. Production of clones of homozygous diploid zebra fish (Brachydanio rerio). Nature 291, 293-296 (1981).
2. Lieschke, G.J. & Currie, P.D. Animal models of human disease: zebrafish swim into view. Nat Rev Genet 8, 353-367 (2007).
3. Brittijn, S.A. et al. Zebrafish development and regeneration: new tools for biomedical research. International Journal of Developmental Biology 53, 835-850 (2009).
4. Ali, S., Champagne, D.L., Spaink, H.P. & Richardson, M.K. Zebrafish embryos and larvae: a new generation of disease models and drug screens. Birth defects research. Part C, Embryo today : reviews 93, 115-133 (2011).
5. Meijer, A.H. & Spaink, H.P. Host-pathogen interactions made transparent with the zebrafish model. Curr.Drug Targets. 12, 1000-1017 (2011).
6. Kari, G., Rodeck, U. & Dicker, A.P. Zebrafish: An Emerging Model System for Human Disease and Drug Discovery. Clin Pharmacol Ther 82, 70-80 (2007).
7. Zebrafish International Resource Center (ZIRC), Universityt of Oregon, USA, (http://zebrafish.org/home/guide.php)
8. Howe, K. et al. The zebrafish reference genome sequence and its relationship to the human genome. Nature advance online publication (2013).
9. Reimers, M.J., Hahn, M.E. & Tanguay, R.L. Two zebrafish alcohol dehydrogenases share common ancestry with mammalian class I, II, IV, and V alcohol dehydrogenase genes but have distinct functional characteristics. J Biol Chem 279, 38303-38312 (2004).
10. Renier, C. et al. Genomic and functional conservation of sedative-hypnotic targets in the zebrafish. Pharmacogenetics and genomics 17, 237-253 (2007).
11. Li, C., Tan, X.F., Lim, T.K., Lin, Q. & Gong, Z. Comprehensive and quantitative proteomic analyses of zebrafish plasma reveals conserved protein profiles between genders and between zebrafish and human. Scientific Reports 6, 24329 (2016).
12. Meijer, A.H. et al. Expression analysis of the Toll-like receptor and TIR domain adaptor families of zebrafish. Molecular immunology 40, 773-783 (2004).
13. Stein, C., Caccamo, M., Laird, G. & Leptin, M. Conservation and divergence of gene families encoding components of innate immune response systems in zebrafish. Genome Biol 8, R251 (2007).
14. Traver, D. et al. The zebrafish as a model organism to study development of the immune system. Adv.Immunol. 81, 253-330 (2003).

15. Trede, N.S., Langenau, D.M., Traver, D., Look, A.T. & Zon, L.I. The use of zebrafish to understand immunity. Immunity. 20, 367-379 (2004).
16. Hasin, Y., Seldin, M. & Lusis, A. Multi-omics approaches to disease. Genome Biology 18, 83 (2017).
17. Overall, C.M. Can proteomics fill the gap between genomics and phenotypes? Journal of Proteomics 100, 1-2 (2014).
18. List, E.O. et al. The use of proteomics to study infectious diseases. Infectious disorders drug targets 8, 31-45 (2008).
19. Plas-Duivesteijn, S.J. et al. Identifying proteins in zebrafish embryos using spectral libraries generated from dissected adult organs and tissues. J Proteome Res 13 (2014).
20. ZFIN. (https://zfin.org/action/genotype/view/ZDB-GENO-960809-7) (2016).
21. Balik-Meisner, M., Truong, L., Scholl, E.H., Tanguay, R.L. & Reif, D.M. Population genetic diversity in zebrafish lines. Mammalian genome : official journal of the International Mammalian Genome Society (2018).
22. Guryev, V. et al. Genetic variation in the zebrafish. Genome Research 16, 491-497 (2006).
23. Carvalho, R. et al. A high-throughput screen for tuberculosis progression. PloS one 6, e16779 (2011).
24. Veneman, W.J. et al. A zebrafish high throughput screening system used for Staphylococcus epidermidis infection marker discovery. BMC Genomics 14, 255 (2013).
25. Pardo-Martin, C. et al. High-throughput in vivo vertebrate screening. Nat Meth 7, 634-636 (2010).
26. Lessman, C.A. The developing zebrafish (Danio rerio): a vertebrate model for high-throughput screening of chemical libraries. Birth defects research. Part C, Embryo today : reviews 93, 268-280 (2011).
27. Letamendia, A. et al. Development and validation of an automated high-throughput system for zebrafish in vivo screenings. PloS one 7, e36690 (2012).
28. Spaink, H.P. et al. Robotic injection of zebrafish embryos for high-throughput screening in disease models. Methods 62, 246-254 (2013).
29. Miscevic, F., Rotstein, O. & Wen, X.Y. Advances in zebrafish high content and high throughput technologies. Comb Chem High Throughput Screen 15, 515-521 (2012).
30. Cui, C. et al. Infectious disease modeling and innate immune function in zebrafish embryos. Methods Cell Biol. 105, 273-308 (2011).
31. Lam, S.H., Chua, H.L., Gong, Z., Lam, T.J. & Sin, Y.M. Development and maturation of the immune system in zebrafish, Danio rerio: a gene expression profiling, in situ hybridization and immunological study. Dev.Comp Immunol. 28, 9-28 (2004).
32. White, R.M. et al. Transparent adult zebrafish as a tool for in vivo transplantation analysis. Cell stem cell 2, 183-189 (2008).
33. ZIRC zebrafish cell lines. (http://zebrafish.org/fish/lineAll.php) (2018)
34. Simonetti, R.B.M., L.S.; Streit, D.P.; Oberst E.R. Zebrafish (Danio rerio): Ethics in animal experimentation. IOSR Journal of Agriculture and Veterinary Science (IOSR-JAVS) 9, 106-110 (2016).
35. Lidster, K., Readman, G.D., Prescott, M.J. & Owen, S.F. International survey on the use and welfare of zebrafish Danio rerio in research. Journal of Fish Biology 90, 1891-1905 (2017).

36. Strahle, U. et al. Zebrafish embryos as an alternative to animal experiments--a commentary on the definition of the onset of protected life stages in animal welfare regulations. Reproductive toxicology (Elmsford, N.Y.) 33, 128-132 (2012).
37. Kimmel, C.B., Ballard, W.W., Kimmel, S.R., Ullmann, B. & Schilling, T.F. Stages of embryonic development of the zebrafish. Developmental dynamics : an official publication of the American Association of Anatomists 203, 253-310 (1995).
38. Adatto, I., Lawrence, C., Thompson, M. & Zon, L.I. A new system for the rapid collection of large numbers of developmentally staged zebrafish embryos. PloS one 6, e21715 (2011).
39. Benard, E.L. et al. Infection of zebrafish embryos with intracellular bacterial pathogens. J.Vis.Exp. (2012).
40. Henry, K.M., Loynes, C.A., Whyte, M.K. & Renshaw, S.A. Zebrafish as a model for the study of neutrophil biology. Journal of leukocyte biology 94, 633-642 (2013).
41. Takaki, K., Cosma, C.L., Troll, M.A. & Ramakrishnan, L. An in vivo platform for rapid high-throughput antitubercular drug discovery. Cell reports 2, 175-184 (2012).
42. Westerfield, M. The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio). Edn. 4. (Univ. of Oregon Press, Eugene; 2000).
43. Link, V., Shevchenko, A. & Heisenberg, C.P. Proteomics of early zebrafish embryos. BMC.Dev.Biol. 6, 1 (2006).
44. Mandrell, D. et al. Automated zebrafish chorion removal and single embryo placement: optimizing throughput of zebrafish developmental toxicity screens. Journal of laboratory automation 17, 66-74 (2012).
45. Choudhury, D. et al. Fish and Chips: a microfluidic perfusion platform for monitoring zebrafish development. Lab on a chip 12, 892-900 (2012).
46. Zheng, C. et al. Fish in chips: an automated microfluidic device to study drug dynamics in vivo using zebrafish embryos. Chemical communications (Cambridge, England) 50, 981-984 (2014).
47. Wang, H., Tan, J.T.T., Emelyanov, A., Korzh, V. & Gong, Z. Hepatic and extrahepatic expression of vitellogenin genes in the zebrafish, Danio rerio. Gene 356, 91-100 (2005).
48. Lin, Y., Chen, Y., Yang, X., Xu, D. & Liang, S. Proteome analysis of a single zebrafish embryo using three different digestion strategies coupled with liquid chromatography-tandem mass spectrometry. Anal.Biochem. 394, 177-185 (2009).
49. Rahlouni, F., Szarka, S., Shulaev, V. & Prokai, L. A Survey of the Impact of Deyolking on Biological Processes Covered by Shotgun Proteomic Analyses of Zebrafish Embryos. Zebrafish 12, 398-407 (2015).
50. IgY RuBisCO columns, (https://www.sigmaaldrich.com/technical-documents/articles/biology/seppro-rubisco.html) (2018)
51. Irmler, I., Schmidt, K. & Starck, J.M. Developmental variability during early embryonic development of zebra fish, Danio rerio. Journal of experimental zoology. Part B, Molecular and developmental evolution 302, 446-457 (2004).
52. Parichy, D.M., Elizondo, M.R., Mills, M.G., Gordon, T.N. & Engeszer, R.E. Normal table of postembryonic zebrafish development: staging by externally visible anatomy of the living fish. Developmental dynamics : an official publication of the American Association of Anatomists 238, 2975-3015 (2009).
53. Bennike, T.B. et al. Comparing the proteome of snap frozen, RNAlater preserved, and formalin-fixed paraffin-embedded human tissue samples. EuPA Open Proteomics 10, 9-18 (2016).

54. Baghirova, S., Hughes, B.G., Hendzel, M.J. & Schulz, R. Sequential fractionation and isolation of subcellular proteins from tissue or cultured cells. MethodsX 2, 440-445 (2015).
55. Havanapan, P.O. & Thongboonkerd, V. Are protease inhibitors required for gel-based proteomics of kidney and urine? J Proteome Res 8, 3109-3117 (2009).
56. Zhang, Y., Fonslow, B.R., Shan, B., Baek, M.-C. & Yates, J.R. Protein Analysis by Shotgun/Bottom-up Proteomics. Chemical Reviews 113, 2343-2394 (2013).
57. Victor, B. et al. Partially sequenced organisms, decoy searches and false discovery rates. J.Proteome Res. 11 (2012).
58. Lam, H. et al. Development and validation of a spectral library searching method for peptide identification from MS/MS. Proteomics. 7, 655-667 (2007).
59. Lam, H. et al. Building consensus spectral libraries for peptide identification in proteomics. Nat.Methods 5, 873-875 (2008).
60. Deutsch, E.W. et al. A guided tour of the Trans-Proteomic Pipeline. Proteomics 10, 1150-1159 (2010).
61. Ebrahimie, E., Fruzangohar, M., Moussavi Nik, S.H. & Newman, M. Gene Ontology-Based Analysis of Zebrafish Omics Data Using the Web Tool Comparative Gene Ontology. Zebrafish 14, 492-494 (2017).
62. Hedges, S.B. The origin and evolution of model organisms. Nat Rev Genet 3, 838-849 (2002).
63. Zuckerkandl, E. The Evolution of Hemoglobin. Sci Am 212 (1965).
64. Zuckerkandl, E. & Pauling, L. Molecules as documents of evolutionary history. J Theor Biol 8 (1965).
65. Welker, F. et al. Ancient proteins resolve the evolutionary history of Darwin's South American ungulates. Nature 522, 81-84 (2015).
66. Procopio, N., Chamberlain, A.T. & Buckley, M. Intra- and Interskeletal Proteome Variations in Fresh and Buried Bones. J Proteome Res 16, 2016-2029 (2017).
67. Jiang, X. et al. Method development of efficient protein extraction in bone tissue for proteome analysis. J Proteome Res 6, 2287-2294 (2007).
68. Buckley, M., Collins, M., Thomas-Oates, J. & Wilson, J.C. Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry. Rapid Commun Mass Spectrom 23 (2009).
69. Orlando, L. et al. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. Nature 499, 74-78 (2013).
70. Sigg, M.A. et al. Evolutionary Proteomics Uncovers Ancient Associations of Cilia with Signaling Pathways. Developmental cell 43, 744-762.e711 (2017).
71. Palmblad, M. & Deelder, A.M. Molecular phylogenetics by direct comparison of tandem mass spectra. Rapid Commun.Mass Spectrom. 26, 728-732 (2012).
72. ATCC FISH CELL LINES, (https://www.lgcstandards-atcc.org/Products/Cells_and_Microorganisms/Cell_Lines/Animal/Fish.aspx?geo_country=nl) (2017).
73. Kolder, I.C.R.M. et al. A full-body transcriptome and proteome resource for the European common carp. BMC Genomics 17, 701 (2016).
74. Smith, S.A. Zebrafish resources on the internet. ILAR journal 53, 208-214 (2012).
75. The UniProt Consortium UniProt: the universal protein knowledgebase. Nucleic Acids Research 45, D158-D169 (2017).
76. Deutsch, E.W., Lam, H. & Aebersold, R. PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. EMBO Rep 9, 429-434 (2008).
77. Peptide atlas: zebrafish (https://db.systemsbiology.net/sbeams/cgi/PeptideAtlas/buildDetails?atlas_build_id=408) (2018)

78. NIST zebrafish database. (http://chemdata.nist.gov/dokuwiki/doku.php?id=peptidew:lib:zebrafish20150109) (2018)
79. Travin, D. et al. COMICS: Cartoon Visualization of Omics Data in Spatial Context Using Anatomical Ontologies. J Proteome Res (2017).
80. Davis, J.M. et al. Real-time visualization of mycobacterium-macrophage interactions leading to initiation of granuloma formation in zebrafish embryos. Immunity. 17, 693-702 (2002).
81. Lesley, R. & Ramakrishnan, L. Insights into early mycobacterial pathogenesis from the zebrafish. Current opinion in microbiology 11, 277-283 (2008).
82. Swaim, L.E. et al. Mycobacterium marinum infection of adult zebrafish causes caseating granulomatous tuberculosis and is moderated by adaptive immunity. Infect.Immun. 74, 6108-6117 (2006).
83. Clay, H. et al. Dichotomous role of the macrophage in early Mycobacterium marinum infection of the zebrafish. Cell host & microbe 2, 29-39 (2007).
84. Tobin, D.M. & Ramakrishnan, L. Comparative pathogenesis of Mycobacterium marinum and Mycobacterium tuberculosis. Cellular microbiology 10, 1027-1039 (2008).
85. Tonjum, T., Welty, D.B., Jantzen, E. & Small, P.L. Differentiation of Mycobacterium ulcerans, M. marinum, and M. haemophilum: mapping of their relationships to M. tuberculosis by fatty acid profile analysis, DNA-DNA hybridization, and 16S rRNA gene sequence analysis. Journal of clinical microbiology 36, 918-925 (1998).
86. Stinear, T.P. et al. Insights from the complete genome sequence of Mycobacterium marinum on the evolution of Mycobacterium tuberculosis. Genome Res 18, 729-741 (2008).
87. van der Vaart, M., Spaink, H.P. & Meijer, A.H. Pathogen recognition and activation of the innate immune response in zebrafish. Advances in hematology 2012, 159807 (2012).
88. Cronan, M.R. & Tobin, D.M. Fit for consumption: zebrafish as a model for tuberculosis. Disease models & mechanisms 7, 777-784 (2014).
89. Volkman, H.E. et al. Tuberculous granuloma induction via interaction of a bacterial secreted protein with host epithelium. Science 327, 466-469 (2010).
90. Volkman, H.E. et al. Tuberculous granuloma formation is enhanced by a mycobacterium virulence determinant. PLoS Biol 2, e367 (2004).
91. Verhoeckx, K.C.M. et al. A combination of proteomics, principal component analysis and transcriptomics is a powerful tool for the identification of biomarkers for macrophage maturation in the U937 cell line. PROTEOMICS 4, 1014-1028 (2004).
92. Hecker, M., Antelmann, H., Buttner, K. & Bernhardt, J. Gel-based proteomics of Gram-positive bacteria: a powerful tool to address physiological questions. Proteomics 8, 4958-4975 (2008).
93. van der Sar, A.M., Spaink, H.P., Zakrzewska, A., Bitter, W. & Meijer, A.H. Specificity of the zebrafish host transcriptome response to acute and chronic mycobacterial infection and the role of innate and adaptive immune components. Mol. Immunol. 46, 2317-2332 (2009).
94. Rougeot, J. et al. RNA Sequencing of FACS-Sorted Immune Cell Populations from Zebrafish Infection Models to Identify Cell Specific Responses to Intracellular Pathogens. Methods Mol Biol 1197, 261-274 (2014).